Addiction (1995) 90, 637–645

RESEARCH REPORT

Factor analysis of alcohol abuse and dependence symptom items in the 1988 National Health Interview survey

BENGT O. MUTHÉN

Graduate School of Education, UCLA, Los Angeles, USA

Abstract

The findings presented in this report are of general interest for the development of survey instruments for alcohol use disorders. They show which items represent the same dimension of alcohol problems when presented to respondents in a general population survey setting. The note determines the major dimensions underlying the complete set of 41 symptom items in the 1988 National Health Interview Survey, relates the item sets of the DSM- and ICD-criteria to these dimensions, and studies the measurement characteristics of items not used for DSM- or ICD-criteria.

Introduction

This report presents factor analysis results for symptom items used in the Alcohol Supplement of the 1988 National Health Interview Survey (NHIS88), a nationally representative, general population survey on alcohol abuse and dependence. These symptom items were specifically developed to capture diagnostic criteria for alcohol abuse and dependence according to definitions given in the *International Classification of Diseases, 10th Revision* (ICD-10) (World Health Organization, 1992), *Diagnostic and Statistical Manual—3rd Edition—Revised* (DSM-III-R) (American Psychiatric Association, 1987), and the proposed DSM-IV (American Psychiatric Association, 1992).

Previous factor analyses of the NHIS88 data have considered subsets of the 41 symptom items. Muthén, Grant & Hasin (1993a,b) presented factor analyses of diagnostic criteria for alcohol abuse and dependence based on symptom items capturing the DSM-III-R and DSM-IV criteria. Muthén, Hasin & Wisnicki (1993b) presented factor analyses of symptom items capturing ICD-10 criteria. It is of interest to take this work further and study the correlational structure among all symptom items used in DSM- or ICD-10 criteria. The relationship with the items not used in such criteria is also of interest. Although theoretically not corresponding to DSM- or ICD-10 criteria, these other items represent an untapped source of information on the status of the individual's alcohol problems.

The aims of this paper are thus to determine the major dimensions underlying the complete set of 41 symptom items in the NHIS88, to relate the item sets of the DSM- and ICD-criteria to these dimensions, to assess the quality with which the items measure these dimensions and to study the measurement characteristics of items not used for DSM- or ICD-criteria. This investigation is of general interest for the development of survey instruments for alcohol use disorders. As shown in Muthén (1995a), there is

Correspondence to: Bengt O. Muthén, Graduate School of Education, UCLA, Los Angeles, CA 90024–1521, USA.

0965–2140/95/050637–09

a need to use many good items to form diagnoses based on survey responses in order to ensure a low degree of misdiagnosis.

Methods

Sample

In 1988, a nationally representative survey on alcohol use and alcohol-related problems was sponsored by the National Institute on Alcohol Abuse and Alcoholism. The data were collected as part of the NHIS88 conducted by the National Center for Health Statistics. The NHIS88 featured a complex, multi-stage sample design that has been described elsewhere (Massey *et al.*, 1989). The design was both stratified and clustered, and blacks were oversampled. The NHIS88 sample consisted of 47 485 households. Health and demographic data were collected for all 122 310 members of those households. For the alcohol supplement, one adult 18 years of age or older was randomly selected from each household, and 43 809 individuals responded. Including non-response to the basic health and demographic component, the overall response rate for the alcohol supplement was 85.5%.

Drinking and high risk subgroups. Of the 43 809 individuals responding to the alcohol supplement, 22 102 were classified as current drinkers by an affirmative answer to the question, "In the past 12 months, did you have at least 12 drinks of any kind of alcoholic beverage?". We present separate analyses for current drinkers and a subgroup of drinkers defined at high risk for alcohol dependence. For the high risk group, individuals were selected based on background characteristics predictive of alcohol dependence. Drawing on analyses discussed in Muthén (1995b), these characteristics included alcohol consumption, family history of alcohol problems, age, gender, ethnicity, marital status, education and income. The individuals with the 12% highest estimated probability of alcohol dependence based on these characteristics were identified ($n = 1164$). This is the percentage of current drinkers classified as alcohol-dependent by DSM-III-R in the NHIS88. Of these high-risk subjects, 49% were alcohol dependent according to DSM-III-R.

Given the large sample sizes, it was possible to split the sample of current drinkers and the sample of the high risk group into two random halves, using one half as the calibration sample to explore models and the other half for validation, minimizing the risk of capitalizing on chance features in the analyses.

Symptom items

The wording of the 41 symptom items is presented in Table 1. The questions were presented to the respondents as "How many times in the past 12 months have you…", with the four response alternatives none or never, one time, two or three times, four or more times. The table also gives the proportion of individuals in the calibration sample of current drinkers reporting the occurrence of each symptom item on two or more occasions in the past year.

Statistical analysis

The set of symptom items was factor analysed as binary 0/1 variables using two alternative dichotomizations. Using the current drinker sample, category 1 corresponds to the symptom having occurred two or more times in the last 12 months. Using the high risk group the category of four or more times was also used since this group reports a sufficient number of occurrences. It is of interest to compare the factor patterns for these two categorizations since they may capture different degrees of severity of the problems involved. In this way, three calibration analyses were carried out: current drinkers using two or more occurrences, high risk group using two or more occurrences, and high risk group using four or more occurrences. Three corresponding validation analyses were also carried out.

Factor analysis of binary variables requires special techniques that use nonlinear regressions of the variables on the continuous, unobserved factor. Due to the nonlinearity of regressions, factor analysis of ordinary Pearson product–moment correlations is not appropriate, but tetrachoric correlations were instead used via the LISCOMP computer program (Muthén, 1978, 1987, 1989). Exploratory factor analysis with Promax rotation allowing for correlated factors (see, e.g. Gorsuch, 1983) were used. The number of factors was determined by looking for sharp breaks in the size of the latent roots of the tetrachoric correlation matrix (i.e. using a scree plot; see, e.g. Gorsuch, 1983). Loadings greater

Table 1. *Symptom items, proportions[1] and loadings[2]*

		Current drinkers (⩾2)		High risk group (⩾2)		High Risk group (⩾4)	
	Proportions	F1	F2	F1	F2	F1	F2
Larger (DSM-III-R, DSM-IV, ICD-10)							
Drinking in larger amounts or over a longer period than the person intended							
3.[3] Ended up drinking much more than you intended to?	0.25	1.05	−0.20	1.06	−0.29	0.95	−0.31
4. Found it difficult to stop drinking once you had started?	0.08	0.66	0.23	0.68	0.18	0.63	0.14
9. Kept on drinking for a longer period of time than you intended to?	0.13	0.82	0.05	0.90	−0.04	0.80	−0.08
2. Started drinking even though you hadn't intended to?	0.17	0.75	0.03	0.91	−0.18	0.85	0.25
Cutdown (DSM-III-R, DSM-IV, ICD-10)							
Persistent desire or one or more unsuccessful efforts to cut down or control							
16. Tried to cut down or stop drinking and found you couldn't do it?	0.02	0.20	0.73	0.32	0.58	0.21	0.70
41. Wanted to cut down or stop your drinking and found you couldn't do it?	0.01	0.19	0.77	0.26	0.66	0.25	0.67
Time spent (DSM-III-R, DSM-IV)							
Spent a great deal of time obtaining alcohol, drinking, or recovering from drinking							
32. Spent a lot of time drinking, or getting over the effects of drinking?	0.03	0.58	0.39	0.62	0.22	0.64	0.22
Majorole (DSM-III-R, DSM-IV)							
Frequent intoxication or withdrawal symptoms when expected to fulfill major role obligations at work, school, or home							
22. Stayed away from work or gone to work late because of drinking or a hangover?	0.03	0.54	0.33	0.44	0.42	0.53	0.26
25. Gotten drunk instead of doing the things you were supposed to do?	0.04	0.57	0.36	0.57	0.26	0.55	0.27
33. Been so hungover that it interferred with doing things you were supposed to do?	0.04	0.66	0.26	0.83	0.06	0.72	0.09
23. Spend money on drinks that was needed for essentials like food or bills?	0.02	0.41	0.52	0.44	0.34	0.44	0.39

Table 1. *Continued*

	Proportions	Current drinkers (⩾2) F1	Current drinkers (⩾2) F2	High risk group (⩾2) F1	High risk group (⩾2) F2	High Risk group (⩾4) F1	High Risk group (⩾4) F2
Hazard (DSM-III-R, DSM-IV)							
Recurrent drinking in situations in which it is physically hazardous							
5. Driven a car after having had too much to drink?	0.10	0.93	−0.14	0.67	0.13	0.69	−0.05
7. Done things when drinking for a longer period of time than you intended to?	0.07	0.86	0.04	0.80	0.05	0.81	−0.02
30. Done things when drinking that could have caused someone else to be hurt?	0.04	0.67	0.20	0.68	0.11	0.63	0.13
Give up (DSM-III-R, DSM-IV, ICD-10)							
Important social, occupational, or recreational activities given up or reduced because of drinking							
18. Given up or cut down on activities or interests like sports or associations with friends, in order to drink?	0.01	0.30	0.67	0.24	0.72	0.27	0.69
24. Lost ties with or drifted apart from a family member or friend because of your drinking?	0.01	0.16	0.82	0.04	0.91	0.10	0.85
Continue (DSM-III-R, DSM-IV, ICD-10)							
Continued to drink depsite knowledge of a persistent or recurrent social, psychological or physical problem that is caused or exacerbated by drinking							
27. Continued to drink alcohol even though it was a threat to your health?	0.02	0.01	0.85	0.12	0.68	−0.06	0.81
34. Kept drinking even though it caused you emotional problem?	0.02	0.32	0.67	0.34	0.62	0.26	0.66
38. Kept drinking, even though it caused you problems at home, work or school?	0.02	0.23	0.74	0.24	0.73	0.27	0.68
14. Had a spouse or someone you lived with threaten to leave because of your drinking?	0.01	0.14	0.73	0.05	0.79	0.03	0.80
28. Lost a job, or nearly lost one, because of your drinking?	0.01	−0.10	0.99	−0.38	1.24	−0.33	1.11
35. Had your chances for promotion, raises or better jobs hurt by yourdrinking?	0.01	−0.07	0.97	−0.16	1.08	−0.15	0.98

Tolerance (DSM-III-R, DSM-IV, ICD-10)							
Tolerance							
10. Found that the same amount of alcohol had less effect than before?	0.09	0.60	0.21	0.59	0.26	0.53	0.16
21. Found that you had to drink more than you once did to get the same effect?	0.04	0.47	0.39	0.43	0.39	0.38	0.32
Withdrawal (DSM-III-R, DSM-IV, ICD-10)							
Characteristic withdrawal symptoms							
6. Been sick or vomited after drinking or the morning after?	0.09	0.75	0.02	0.66	−0.01	0.60	−0.04
11. Felt depressed, irritable or nervous after drinking or the morning after?	0.10	0.62	0.25	0.44	0.45	0.46	0.33
36. Heard or seen things that weren't really there after drinking or the morning after?	0.01	0.17	0.74	−0.03	0.94	0.11	0.73
17. Found yourself sweating heavily or shaking after drinking or the morning after?	0.01	0.45	0.45	0.55	0.29	0.43	0.38
Relief (DSM-III-R, DSM-IV, ICD-10)							
Drinking to relieve or avoid withdrawal symptoms							
37. Taken a drink to keep yourself from shaking or feeling sick either after drinking or the morning after?	0.03	0.27	0.66	0.34	0.56	0.36	0.51
Legal (DSM-IV)							
Recurrent alcohol-related legal or interpersonal problems							
40. Been arrested or had trouble with the police because of your drinking?	0.01	0.27	0.56	0.30	0.46	0.32	0.40
Compulsion (ICD-10)							
A strong desire or sense of compulsion to drink							
1. Had a strong desire or urge to drink?	0.19	0.69	0.10	0.93	−0.23	0.65	0.03
12. Felt powerless over your drinking?	0.02	0.30	0.65	0.36	0.56	0.24	0.67
20. Needed a drink so badly you couldn't think of anything else?	0.01	0.38	0.60	0.43	0.47	0.35	0.56

Table 1. *Continued*

			Current drinkers (⩾2)		High risk group (⩾2)		High Risk group (⩾4)	
		Proportions	F1	F2	F1	F2	F1	F2
Other								
8.	Felt the effects of alcohol sooner than you used to?	0.13	0.52	0.12	0.72	−0.07	0.54	0.01
13.	Sought help from family, friends, professionals or self-help groups about your drinking?	0.01	0.15	0.79	0.25	0.63	0.19	0.72
15.	Gone on benders or binges that lasted two or more days?	0.02	0.39	0.52	0.23	0.58	0.31	0.48
19.	Been unable to remember some of the things you did while drinking?	0.07	0.66	0.25	0.62	0.20	0.62	0.17
26.	Had a doctor suggest that you cut down or stop drinking alcohol?	0.01	−0.01	0.87	0.06	0.88	0.01	0.84
29.	Had family, friends or co-workers suggest that you stop or cut down on your drinking?	0.03	0.11	0.78	0.16	0.66	0.10	0.73
31.	Felt uneasy if alcohol was not around in case you wanted a drink?	0.02	0.25	0.64	0.35	0.52	0.26	0.55
39.	Attended a meeting of Alcoholics Anonymous (AA) because of your drinking?	0.01	0.04	0.84	0.21	0.63	0.11	0.71

[1] Proportion of individuals in the calibration sample of current drinkers reporting the occurrence of a symptom item on two or more occasions (⩾2).
[2] Factor loadings from pattern matrix.
[3] Numbering refers to the order in which the items were presented in the interview.

than one were allowed in order to indicate a "Heywood case", i.e. a case where the residual variance obtains a negative estimate, usually due to large sampling variance.

Findings

For all three calibration factor analyses a two-factor solution was suggested by the size of the latent roots. In all cases, the first root was much larger than the rest, the second root was about twice as big as the third, while the third and consecutive roots showed considerably smaller differences. The analyses of the validation samples gave similar results in this regard and also gave very similar two-factor solutions.

The three calibration factor analyses are given in Table 1 for current drinkers and the high risk group with the two alternative categorizations. The items of the table are organized according to the DSM-III-R, DSM-IV and ICD-10 criteria, using the shorthand notation Larger–Compulsion. Under each shorthand notation is given the definition of the criterion. The group of items labeled Other refers to items not used in the NHIS88 to form DSM-III-R, DSM-IV or ICD-10 criteria. Table 1 shows that the factor loading patterns are very similar across these three analyses. The two factors are rather highly correlated, 0.72, 0.70 and 0.65, respectively, for the three analyses. The factor pattern can be described as follows.[1] Factor 1 (F1) is well measured by the items of the criteria Larger and Hazard. Factor 2 (F2) is well measured by the items of the criteria Cutdown, Giveup, Continue, Relief and Legal. Judging by the corresponding items, F1 is represented by more prevalent symptoms than F2 and defines a dimension of less severe problems than F2.

For most criteria, all items load on one and not the other factor, indicating that the corresponding set of items is homogeneous and measures its factor well. Exceptions are the Majorole item 23, the Tolerance item 21, the Withdrawal items and the Compulsion item 1. For the Other items, i.e. the items not used in the DSM- or ICD-criteria, every item except item 15 loads strongly on one but not the other factor.

[1] For researchers wanting to replicate these findings it should be noted that to simplify the presentation, the order of the factors is reversed relative to what is given in the computer output. The order is arbitrary in factor analysis given the rotational indeterminacy.

Conclusion

The two-factor solution presented above bears strong resemblance to that presented in Muthén, Grant & Hasin (1993a), analysing NHIS88 data on the 11 DSM criteria for abuse and dependence shown in Table 1. The first factor corresponds to less severe and more prevalent symptom items than the second factor. As in the present analysis, the less severe factor was represented by the criteria Larger and Hazard, while the more severe factor was represented by the criteria Cutdown, Continue and Giveup. The criteria Larger and Hazard are considerably more prevalent than the other criteria in this general population survey. In line with Muthén *et al.* (1993a), the less severe factor that they measure may be seen as corresponding to abuse, while the more severe factor may be seen as a dependence factor. Muthén, Hasin & Wisnicki (1991) recovered the same major dimensions for the items of the ICD-10 criteria.

It is interesting to note that the item-level two-factor solution remains the same across the three analyses of Table 1. In terms of factor pattern, it does not matter whether one considers a higher-risk group or restricts attention to more frequent occurrences of the symptoms. The two dimensions are defined in the same way and have about the same correlation in all three cases.

Most criteria have items that measure only one of the two factors. For such criteria this is an indication that the survey items are homogeneous, measuring the criterion well. This is a fortunate situation where the theoretical intentions behind the creation of the survey item correspond to how the survey respondents react to the item in the interview setting. Such good measurement properties are observed for the items of the F1 (abuse) criteria Larger and Hazard, as well as for the F2 (dependence) criteria Cutdown, Giveup and Continue.

A notable exception to item homogeneity is item 1, "Had a strong desire or urge to drink?", of the Compulsion criterion. This criterion is used in ICD-10 and is intended to capture a central aspect of alcohol dependence. As expected, the other two Compulsion items load strongly on only the more severe factor F2, but item 1 loads on F1. This picture does not change when considering the results from the high risk group or when considering a cutoff at four or more occurrences. The same finding is made in the validation sample. The proportions in Table

1 show that item 1 is considerably more prevalent than the other two items. As understood by the respondents in this survey, this item therefore does not by itself function as an indicator of the alcohol dependence dimension.

For some criteria the items show unexpected loading patterns. The Majorole items load mostly on F1 which is unexpected, since this is the less severe factor, and Majorole is aimed at capturing failures of fulfilling major role obligations. Similarly, the items of the Tolerance and Withdrawal criteria load in several cases on F1. For Withdrawal it is only the rare item 36, "Heard or seen things that weren't really there after drinking or the morning after" that clearly loads on the more severe factor F2. The same findings are made in the validation sample. This may be an indication that these criteria are difficult to operationalize in survey settings. However, the fact that individual symptom items do not represent the more severe factor F2 does not preclude that the factor is well measured by criteria based on these items. In the NHIS88 a diagnostic criterion is, in general, considered to be fulfilled if at least one of its symptoms was experienced at least twice in the last year. Given the central importance of Tolerance and Withdrawal, however, two and three items, respectively, were required for these criteria to be fulfilled. As shown in different factor solutions in Muthén, Grant & Hasin (1993a,b) and in Muthén (1995b) the requirement of joint occurrence of several symptom items allows the criteria to reflect the more severe factor.

The factor analyses suggest that two dimensions can explain important parts of the correlations among the alcohol symptom items. Given this finding, it is interesting to note that the eight items of the Other category, i.e. those symptom items not used in DSM- or ICD-criteria, also measure these two dimensions well. Excellent indicators of the more severe dimension F2 include items 26 and 29 where a doctor, family, friends or co-workers suggest that the person stop or cut down on the drinking. This more severe dimension is also well indicated by items 13 and 39, having sought help about drinking, for example from Alcoholics Anonymous. If the aim is to measure the two dimensions well, as opposed to measuring specific diagnostic criteria, these Other items are very useful.

The findings presented in this Data Note are of general interest for the development of survey instruments for alcohol use disorders. Table 1 shows which items represent the same dimension of alcohol problems when presented to respondents in a general population survey setting. The results give ideas about which diagnostic criteria are relatively easy to operationalize and which are relatively difficult to capture. They indicate areas of alcohol problems for which it is profitable to develop more good survey items.

Acknowledgements

This research was supported by grant AA 08651-01 from NIAAA for the project "Psychometric Advances for Alcohol and Depression Studies". Jin-Wen Hsu, Ginger Nelson, Li-Chiao Huang and Siek-Toon Khoo provided helpful research assistance. I thank Dr Bridget Grant for comments on an earlier version.

References

AMERICAN PSYCHIATRIC ASSOCIATION (1987) *Diagnostic and Statistical Manual of Mental Disorders*, 3rd edn, rev. (DSM-III-R) (Washington, DC, American Psychiatric Association).

AMERICAN PSYCHIATRIC ASSOCIATION (1992) *DSM-IV Options Book* (Washington, DC, American Psychiatric Association).

GORSUCH, R. L. (1983) *Factor Analysis*, 2nd edn (Hillsdale, NJ, Lawrence Erlbaum Associates).

MASSEY, J. T., MOORE, T. F., PARSONS, V. L. & TADORS, W. (1989) Design and estimation for the National Health Interview Survey, 1985–1994, *Vital and Health Statistics*, 2, 1–33.

MUTHÉN, B. (1978) Contributions to factor analysis of dichotomous variables, *Psychometrika*, 43, 551–560.

MUTHÉN, B. (1987) LISCOMP. *Analysis of Linear Structural Equations with a Comprehensive Measurement Model, Theoretical integration and user's guide* (Mooresville, IN, Scientific Software).

MUTHÉN, B. (1989) Dichotomous factor analysis of symptom data, in: EATON, W. W. & BOHRNSTEDT, G. (eds) Latent Variable Models for Dichotomous Outcomes: Analysis of Data from the Epidemiological Catchment Area Program, *Sociological Methods & Research*, 18, 19–65.

MUTHÉN, B. (1995a) Psychometric evaluation of measurement properties in a two-dimensional model of alcohol abuse and dependence (manuscript submitted).

MUTHÉN, B. (1995b) Covariates of alcohol dependence and abuse: a multivariate analysis of a 1988 General Population Survey in the United States, *Acta Psychiatrica Scandinavica*, accepted manuscript.

MUTHÉN. B., GRANT, B. & HASIN, D. (1993) The dimensionality of alcohol abuse and dependence:

Factor analysis of DSM-III-R and proposed DSM-IV criteria in the 1988 National Health Interview Survey. *Addiction*, 88, 1079–1090.

MUTHÉN, B., HASIN, D. & WISNICKI, K. (1993b) Factor analysis of ICD-10 symptom items in the 1988 National Health Interview Survey on Alcohol dependence. *Addiction*, 88, 1071–1077.

MUTHÉN, B., GRANT, B. & HASIN, D. (1995) Subgroup differences in factor structure for DSM-III-R and proposed DSM-IV criteria for alcohol abuse and dependence in the 1988 National Health Interview Survey. *Journal of Nervous and Mental Disease*, accepted manuscript.

WORLD HEALTH ORGANIZATION (1992) *International Classification of Diseases—Tenth Revision: Clinical Descriptions and Diagnostic Guidelines* (Geneva, World Health Organization).